

ASSESSING THE POPULATION GENETICS OF *QUERCUS TARDIFOLIA* AND OTHER RED OAKS IN WEST TEXAS

Austin C. Koontz,^{1,*} Silvia Alvarez-Claire,[†] Adam Black,[‡] Amy Byrne,[†] Michael Eason,[§] Emily Griswold,^{||} Marlene Hahn,^{*} Megan Haidet,[#] Wesley Knapp,^{**} Andrew Loyd,[‡] Ray Mims,[#] John Nix,^{§§} Ryan Russell,^{|||} John Saltiel,^{##} Annaliese Sander,[‡] Phillip Schulze,^{***} Robinson Sudan,^{†††} Elizabeth Thomas,[‡] Murphy Westwood,^{*} James Winn VI,[‡] Kelsey Wogan,[‡] Zarah Wylie,^{*} Sean M. Hoban,^{2,*} and Andrew L. Hipp^{2,*}

^{*}Center for Tree Science, Morton Arboretum, Lisle, Illinois, USA; [†]Global Tree Conservation Program, Morton Arboretum, Lisle, Illinois, USA; [‡]Bartlett Tree Research Laboratories and Arboretum, Charlotte, North Carolina, USA; [§]San Antonio Botanical Garden, San Antonio, Texas, USA; ^{||}UC Davis Arboretum and Public Garden, Davis, California, USA; [#]United States Botanic Gardens, Washington, DC, USA; ^{**}Center for Plant Conservation, Wildomar, California, USA; ^{††}NatureServe, Arlington, Virginia, USA; ^{‡‡}Botanical Research Institute of Texas, Fort Worth, Texas, USA; ^{§§}North Carolina State University, Raleigh, North Carolina, USA; ^{|||}Stephen's Lake Park Arboretum, Columbia, Missouri, USA; ^{##}Smithsonian Gardens, Smithsonian Institution, Washington, DC, USA; ^{***}Lady Bird Johnson Wildflower Center, Austin, Texas, USA; ^{†††}Texas State University, San Marcos, Texas, USA; ^{‡‡‡}Sul Ross State University, Alpine, Texas, USA; and ^{||||}Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois, USA

Editor: Jeremie Fant

Premise of research. Successful conservation efforts require clearly delineated species in order to direct resources to conserve threatened populations. Hybridization and admixture can complicate efforts to determine species boundaries, and hybrid individuals are often regarded as abnormal or unimportant in a conservation context. This is despite many examples demonstrating that hybridization can be crucial for adaptation, especially within plants. Here, we describe our research delineating species boundaries in a group of red oaks found in the Chisos and Davis Mountains of west Texas, with a focus on *Quercus tardifolia*, the lateleaf oak, which was thought to have gone extinct in 2011 before its identification in Big Bend National Park in 2022 and other locations thereafter.

Methodology. Restriction site-associated DNA sequencing was used to genotype 166 red oaks, with sampling focused on *Q. tardifolia* and other potentially interbreeding species. Admixed individuals were identified using two different clustering approaches, and phylogenetic networks were used to model past hybridization events and provide validation that admixed individuals are in fact hybrids. We used comparisons of heterozygosity and the NewHybrids software as a further validation of hybridization and conducted flow cytometric analyses to assess ploidy levels in different lineages.

Pivotal results. Our results across analyses consistently inferred all sampled wild *Q. tardifolia* individuals to be hybrids of lineages composed primarily of *Q. gravesii* and *Q. hypoleucoides*, with genetic patterns most resembling those of an F₁ hybrid class. Admixture was also observed in other taxa, but more extensive sampling efforts are required to better characterize the relationships in this group of red oaks.

Conclusions. While our findings indicate that *Q. tardifolia* is likely not a distinct species, more research is needed to determine its true abundance on the landscape and its importance in interspecific introgression or adaptation within this species complex.

Keywords: *Quercus*, Texas, hybridization, population genetics.

Online enhancement: appendix.

¹ Author for correspondence; email: akoontz@mortonarb.org.

² Co-senior authorship.

Manuscript received February 2026; revised manuscript received July 2026; electronically published October 22, 2026.

Introduction

Clear species delimitation aids effective conservation by allowing for the assessment of threatened populations and the appropriate allocation of limited resources (Mace 2004). Threat assessment approaches, legal frameworks, and public perception of biodiversity often operate at the species level (Piett et al. 2015; Noss et al. 2021). However, hybridization and introgression can make it difficult to determine species boundaries, thereby challenging conservation efforts (Stanton et al. 2019; Parmesan et al. 2023). This is especially true when species are only weakly diverged or when gene flow from other species shapes genomic, morphological, or ecological diversity, as is common in many plant species (Mallet 2005), and notoriously so in the genus *Quercus* L. (Trelease 1917; Kremer and Hipp 2020).

Oaks have long been a model group for examining hybridization and the nature of species (MacDougal 1907; Palmer 1948; Tucker 1961; Burger 1975; Petit et al. 2004; Curtu et al. 2007) and consequently have been valuable for interrogating assumptions surrounding species boundaries and gene flow (McVay et al. 2017b). They are also highly phenotypically variable within species, as a result of both genetic variation and plastic variation, further compounding taxonomic challenges (Darwin 1872; Cowan and Baranski 1976; Rubio De Casas et al. 2007; Desmond et al. 2021). Beyond their relevance for understanding hybridization and speciation, oaks are significant ecologically (Cavender-Bares 2019), economically (Cavender-Bares et al. 2022; Frankel et al. 2022), and culturally (Leroy et al. 2020). Since approximately 30% of oak species are threatened with extinction (Carrero et al. 2020), they serve as an important group for understanding how to implement effective conservation despite unclear species boundaries. In the face of oaks' variable morphology and interspecies gene flow, genomic approaches have emerged as a critical tool for supporting or refining oak taxonomy and informing conservation priorities.

In west Texas, taxonomic descriptions of *Quercus* have noted the “extreme variation” in morphology that has made it “difficult, if not impossible” to identify species (Muller 1940, p. 703). In 2019, the Conservation Gap Analysis for US Oaks (Beckman et al. 2019) noted that the three US oaks of greatest conservation concern—*Q. graciliformis* C.H. Mull., *Q. robusta* C.H. Mull., and *Q. tardifolia* C.H. Mull.—are all located in this oak biodiversity hot spot (fig. S1 [figs. S1–S12 are available online]). Reflecting the ambiguity first observed by C. H. Muller and others, the gap analysis report noted “the lack of an articulate, comprehensive taxonomic study” (p. 188) as one of the greatest barriers to conservation of these rare species and that *Q. tardifolia* in particular “has serious taxonomic questions that need to be addressed to confirm its status as a true species” (p. 28). *Quercus tardifolia* was described from the type collection in 1936 and, through 1951, was known only from the type locality, making it difficult to assess its status (Muller 1951). Moreover, *Q. tardifolia* has been considered to be a possible hybrid between *Q. gravesii* Sudw. and *Q. hypoxantha* Trel., which in turn is closely related to *Q. miquihuanensis* Nixon & C.H.Mull. (Jensen 1997). Thus, the genetic status of the species is in question but has never been tested using molecular data.

Quercus tardifolia was thought to have gone extinct in 2011 (Knapp et al. 2021; Renault 2021). However, a targeted

expedition in 2022 recovered single living individual in Big Bend National Park matching the description of *Q. tardifolia* and quite close to the type locality (Renault 2022). Additional putative *Q. tardifolia* individuals were located during subsequent field surveys (September 2022, March 2023, and October 2023) of the Davis Mountains of west Texas. These discoveries allow for a broader assessment of the genetic characteristics of *Q. tardifolia* and an understanding of how *Q. tardifolia* relates to more abundant Texas red oaks (such as *Q. gravesii*, *Q. emoryi*, and *Q. hypoleucoides*; see fig. 1).

As genetic data have become more widely available from wild populations (DeSalle and Amato 2004; Andrews et al. 2016), several classes of techniques based on different underlying models have been developed for resolving species boundaries and characterizing hybrid individuals. A classical method is to infer hybrid individuals based on elevated levels of heterozygosity compared to parental lineages (Barton and Hewitt 1985; Harrison 1986), an approach that remains useful today (Fitzpatrick 2012; Muniz et al. 2020), particularly for early-generation hybrids. More recently developed clustering techniques relying on allele frequencies (Pritchard et al. 2000; Falush et al. 2003; Rosenberg et al. 2003; Alexander et al. 2009) and multivariate, model-free approaches (Patterson et al. 2006; Novembre and Stephens 2008; Ma and Amos 2012) that group individuals according to their genetic similarity. Other approaches use model-based methods to assign individuals to hybrid classes (e.g., F_1 , F_2 , backcross) based on their genotypes (Anderson and Thompson 2002), which is valuable for understanding the origins of introgression and hybrid fertility. Finally, phylogenetic approaches based on multispecies network coalescent models estimate species-level evolutionary history and inheritance probabilities in hybridized lineages (Degnan 2018; Wen et al. 2018).

These types of genetic assessments can determine whether taxonomic divisions between species align with genomic patterns, which can clarify aspects of hybrid individuals, such as their hybrid class (Payseur and Rieseberg 2016; Hausdorf 2025). The development and proliferation of the approaches described above have renewed discussion of species concepts and which criteria are necessary for species delimitation (De Queiroz 2007; Rannala 2015; Huang 2020). While there is no universally adopted set of criteria for delimiting species (Mayden 1997), species concepts based on genotypic groups (Mallet 1995, 2020) are well suited for the clustering approaches described above and accommodate the recognition of distinct species despite continuous gene flow in sympatry (Burger 1975). Crucially, the utilization of multiple approaches based on different underlying assumptions and models is necessary for validating findings and classifying species confidently (Carstens et al. 2013).

An additional important consideration in the assessment of hybridization is polyploidy, since the ploidy levels of hybrid individuals may partially explain their origins or persistence on the landscape. While polyploidy has played a central role in hybridization and species diversification in plants generally (Soltis and Soltis 2009; Alix et al. 2017), there are no known polyploid oak species. Isolated reports of polyploidy in *Quercus* are documented (Burda and Shchepotiev 1973; Butorina 1993), with the frequency of triploids in wild populations reported to be less than 1% (Dzialuk et al. 2007), but these have largely centered on European species

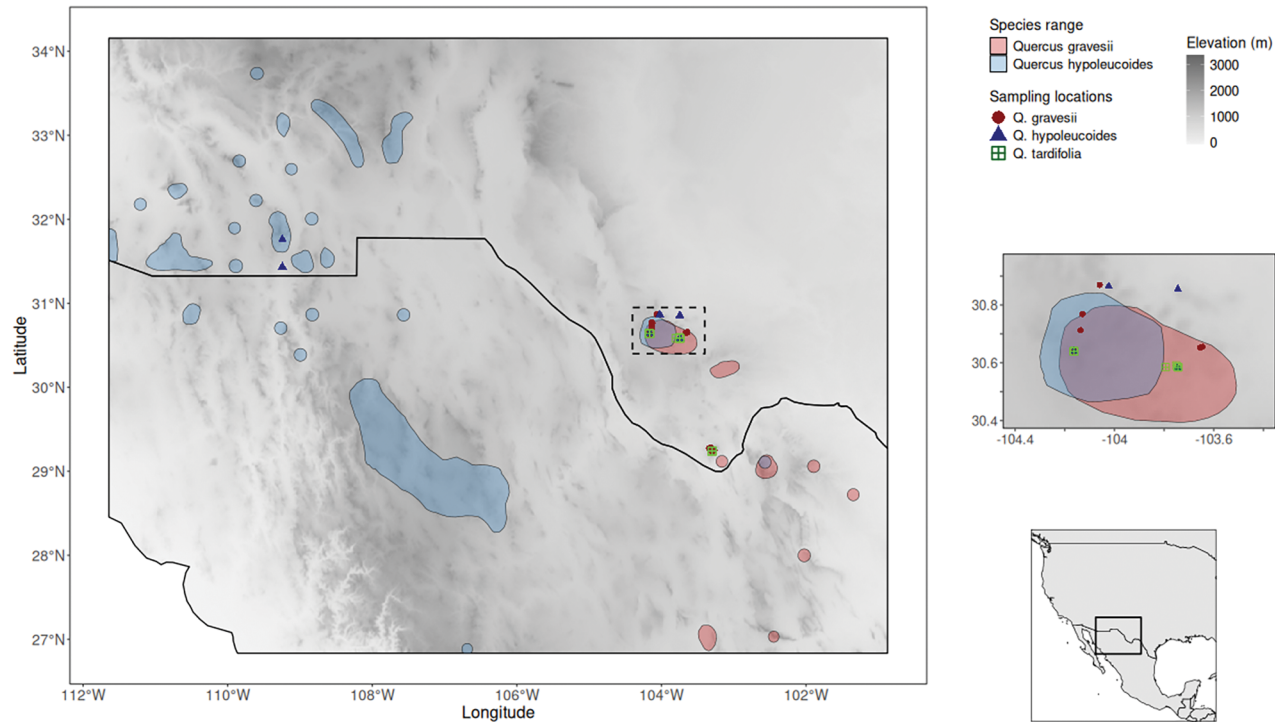


Fig. 1 Partial range maps for *Quercus gravesii* and *Q. hypoleucoides*, with sampling localities for individuals from both species as well as *Q. tardifolia*. Sampling locations of known *Q. gravesii* × *hypoleucoides* individuals not included. The ranges for *Q. gravesii* and *Q. hypoleucoides* were taken from shapefiles (Petty and Taylor 2025) based on ranges described in Little (1971). Note that the ranges shown here are partial and do not represent the entire known extent of either species.

(*Q. robur* L. and *Q. petraea* Matt. Liebl.) and have not been found to coincide with interspecific hybridization. Diploidy and chromosome number are widely conserved across oaks (Sax 1930) and are rarely examined in oak hybridization studies, but the potential for polyploid hybrid lineages within the genus could carry significant implications, especially given increased reproductive isolation between a polyploid hybrid and its parental species.

Here, we use restriction site-associated DNA sequencing (RADseq) data and a range of different genetic analyses to clarify the species boundaries in west Texas red oaks (*Quercus* sect. *Lobatae* Loud.), following the genomic cluster species concept (Mallet 2020) and with a particular emphasis on *Q. tardifolia*. We evaluate how *Q. tardifolia* is related to other taxa on the landscape, whether it is a genetically distinct taxonomic entity, and which taxa are contributing to the genetic makeup of *Q. tardifolia* if it is found to be admixed. By providing the first population genetic assessment of the west Texas red oaks, we aim to disentangle the relations in this area of potentially many interbreeding species and inform future conservation efforts of the rare taxa included in these analyses.

Material and Methods

Sample Collection

Samples were collected on multiple sampling trips from 2021 to 2023. Taxa were chosen for analysis based on which

species are sympatric with *Quercus tardifolia*; additional sequencing data from species of interest or species presumed to be closely related to *Q. tardifolia* (Hipp et al. 2020) were also included. Taxa were included to test for interspecific gene flow involving co-occurring red oak species (*Quercus* sect. *Lobatae*), which are all potential hybridizers, and to assess the genomic coherence of the most significant co-occurring species and close relatives—*Q. gravesii*, *Q. emoryi* Torr., *Q. hypoleucoides* A. Camus, and *Q. scytophylla* Liebm.—across much of their range, including both Arizonan and Mexican populations of those species where possible. Finally, we sought to test the hypothesis that *Q. tardifolia* is a hybrid between *Q. gravesii* and *Q. hypoxantha*.

Contemporary samples were collected to document the morphological, ecological, and geographic variation throughout the range of *Q. tardifolia* as currently understood (Jensen 1997). Timed, meandering surveys were conducted on both public and private lands based on expert review and on-the-ground knowledge of species occurrences, supplemented by review of specimens (Sul Ross State University herbarium [SRSC]).

Our complete sample set was ultimately composed of 166 individuals and one technical replicate, including 10 individuals classified as *Q. tardifolia* or *Q. aff. tardifolia*, based on morphology; individuals from *Q. gravesii*, *Q. graciliformis*, *Q. emoryi*, *Q. miquihuanensis*, and others; and several samples of questionable identity, including putative hybrids (table A1 [tables A1

and A2 are available online]), among them seven samples identified as *Q. gravesii* × *hypoleucoides* (Muller 1951; syn. *Q. × inconstans* E.J. Palmer; see fig. S2). Our sample set also included four herbarium specimens: three of *Q. tardifolia*, including the type specimen, and one of a putative *Q. hypoxantha* × *gravesii*.

Library Preparation and Sequencing

DNA was extracted from fresh, frozen, and herbarium material using the DNeasy plant extraction protocol (DNeasy, Qiagen, Valencia, CA), with modifications previously used in other oak studies (Hipp and Weber 2008; Mohn et al. 2026). DNA concentrations were determined using the Qubit dsDNA HS assay kit. All extractions were diluted to a final concentration of 10 ng/μL (100 μL, total DNA 1 μg). Diluted extractions were visualized in 1.5% agarose gels using Invitrogen 1 Kb Plus DNA ladder as the standard to estimate DNA fragment length distributions of each sample. Following quantification, DNA extractions were sent to Floragenex for final library preparation and sequencing following Baird et al. (2008). All samples were digested using the *PstI* restriction enzyme, and all new material was sequenced on an Illumina NovaSeq 6000.

Genotyping

Genotyping was conducted using Stacks 2 (ver. 2.65; hereafter, “Stacks”; Rivera-Colón and Catchen 2022), and subsequent bioinformatic analyses were run using packages within the R environment (ver. 4.2.2; R Core Team 2022) and other software, as described below. Stacks is a commonly used genotyping software in population genetic analyses and was chosen to allow for control over filtering parameterization and particularly different levels of missing data, which can significantly impact the results of genetic analyses (Hodel et al. 2017). Other software used in analyses is cited accordingly; all scripts used for analyses are available on GitHub (https://github.com/HobanLab/QUTA_TxRedOaks), and releases of the code are versioned and archived on Zenodo (<https://doi.org/10.5281/zenodo.18306919>).

Raw demultiplexed reads were processed in FastQC (Andrews 2010) and Stacks to assess sequencing quality. The process_radtags script in Stacks was used to remove reads with uncalled bases and discard reads with low-quality scores. After quality checking, reads were aligned to the *Q. rubra* reference genome (ver. 2.1; Kapoor et al. 2023) using BWA-MEM 0.7.17 (Li and Durbin 2009), which has been shown to perform well compared to other aligners for read lengths under 500 bp (Musich et al. 2021). Use of the *Q. rubra* reference has the added benefit of minimizing mapping bias, as the species is closely related to but not conspecific with any of our ingroup samples (Mohn et al. 2026). The Stacks ref_map.pl script was then used to build libraries based on reference alignments.

The Stacks populations script was run multiple times to generate different subsets of samples for different downstream analyses. An initial run using all samples, an -R parameter (minimum percentage of individuals in which a locus must be present to be included in the dataset) of 97%, and a minor

allele frequency threshold of 1% was used to evaluate the extent of sequencing error based on the replicate sample (Mastretta-Yanes et al. 2015), identify samples with poor coverage, and assess general clustering patterns (see “Clustering”). Missing data filters for subsequent runs of the populations script are specified below. For all runs of the populations script, no groupings were specified within the popmap file (i.e., all samples were specified as belonging to the same “population”), and only the first single-nucleotide polymorphism (SNP) per locus was used (–write-single-SNP).

Clustering

We used the software STRUCTURE (Pritchard et al. 2000) on various subsets of our data as a starting point for determining how genetically distinct *Q. tardifolia* individuals were from other west Texas red oaks. For all STRUCTURE runs, 20 replicates per *K* value were specified, and the burn-in and number of Markov chain Monte Carlo replicates were each set to 100,000 according to Gilbert et al. (2012). The default admixture (0, modeling admixture) and allele frequency models ($\lambda = 1$) were used, as were default values for all other parameters. The software CLUMPAK (Kopelman et al. 2015) was used to summarize results across replicate STRUCTURE runs and generate CLUMPP files (Jakobsson and Rosenberg 2007), which were visualized using a custom script in R. For each STRUCTURE run, the “Best *K*” module in CLUMPAK was run to determine optimal *K* values using two approaches: the Evanno method (Evanno et al. 2005) and by determining the *K* value with the greatest likelihood (as described in the STRUCTURE manual; Pritchard et al. 2010).

A preliminary STRUCTURE run of the complete dataset (dataset 1) was used to assess general clustering patterns, while a run of a subset of 84 individuals (dataset 2; see “Clustering Analyses” column in table A1) was used to highlight clustering patterns within *Q. tardifolia* (based off of preliminary results in the complete dataset) and to better control for unbalanced representation of different taxa in the complete dataset (Puechmaille 2016; Meirmans 2019). A STRUCTURE run including solely the individuals used for hybrid class assignment and heterozygosity comparisons was also performed (dataset 3; see “Hybrid vs. Parental Comparisons” column in table A1), to further validate signatures of hybridization in this set of individuals. Dataset 1, including all individuals, explored *K* values from *K* = 2 to *K* = 11 and used a missing data filter of R97, leading to 12,123 loci. Dataset 2 explored *K* values from *K* = 2 to *K* = 7 and used a missing data filter of R98, leading to 9954 loci available. Dataset 3 explored *K* values from *K* = 2 to *K* = 7 and used a missing data filter of R98, leading to 23,740 loci available, of which 10,000 were randomly selected and included in analyses (for computational feasibility). All missing data filter values were chosen to minimize the level of missing data while still retaining greater than 5000 loci.

To validate our STRUCTURE results, STRUCTURE datasets 2 and 3 were also processed using principal component analysis (PCA), a model-free approach that makes no assumptions regarding Hardy-Weinberg equilibrium or linkage disequilibrium. The adegenet package (ver. 2.1.10; Jombart 2008) was used to read in Stacks output files and run the PCA. Additionally, we calculated F_{ST} (Weir and Cockerham

1984) to provide a quantitative estimate of the differentiation between the dominant taxa in our analyses. We selected four taxa (*Q. gravesii*, *Q. hypoleucoides*, *Q. tardifolia*, and *Q. emoryi*) based on their prevalence in our complete sample set and the clusters identified in our PCA analysis of dataset 2 (see “ F_{ST} Calculations” column in table A1). The Stacks populations module was used to generate a genepop object (using a missing data filter of R98, leading to 7317 loci), which was passed onto the dartR package (ver. 2.9.7; Gruber et al. 2018) to calculate pairwise F_{ST} values between groups (2000 bootstraps).

Hybridization Analyses

Following our clustering analyses, we sought to better explain patterns of hybridization within *Q. tardifolia* individuals seen in STRUCTURE and PCA results. Even after correcting for uneven sampling, STRUCTURE and similar approaches can represent groups made up of relatively few individuals as admixed, with ancestry coming from other larger groups subject to greater genetic drift (Lawson et al. 2018). To minimize the risk that small populations were being falsely inferred as admixed, we used a phylogenetic network method, comparisons of heterozygosity, and assignment of individuals into hybrid classes to validate signatures of hybridization in our dataset, as described below.

Phylogenetic network. The program SNaQ (Species Networks applying Quartets; Solís-Lemus and Ané 2016) was used to build a phylogenetic network and corroborate signs of hybridization in our dataset using a multispecies network coalescent model. A phylogenetic network is a structure similar to a phylogenetic tree but allows for modeling the transfer of genetic material between distinct lineages through processes such as hybridization and introgression.

To generate a phylogenetic network, a subset of the individuals included in our STRUCTURE run were selected to represent five different groups, based on the clusters observed in our STRUCTURE results: *gravesii*, *hypoleucoides* and *scytophylla*, *emoryi*, *tardifolia*, and an outgroup including *hypoxantha* and *miquihuaensis* individuals. Seven individuals classified as *Q. tardifolia* or *Q. aff. tardifolia* were included in the *Q. tardifolia* group; for the other four groups, the least admixed individuals according to our STRUCTURE analysis of dataset 2 were selected (using a Q value threshold of 0.9 at $K = 4$; see “Phylogenetic Network” column in table A1). A variant call format (VCF) file including these individuals was generated using the Stacks populations script. This VCF file was converted into a table of concordance factors (values quantifying the proportion of gene trees with identical topologies as the species tree) using the SNPs2CF R function (Olave and Meyer 2020). Concordance factors were provided to SNaQ to generate a phylogenetic network and estimate ancestry percentages for modeled hybrid lineages. Networks were inferred using three different values for the maximum number of allowed hybridization events (zero, one, and two), and the suitability of different networks was assessed using log pseudolikelihood scores.

Heterozygosity comparisons. Given signatures of hybridization, we sought to evaluate whether *Q. tardifolia* individuals had statistically higher observed heterozygosity (H_O) compared

to putative parental taxa (*Q. gravesii* and *Q. hypoleucoides*), as expected in F_1 hybrids (Fitzpatrick 2012). To do so, the dartR package was used to read in a VCF file and calculate H_O values in *Q. tardifolia* individuals and groups composed of parental individuals. Seven individuals classified as *Q. tardifolia* (or *Q. aff. tardifolia*) were included in the *Q. tardifolia* group; for *Q. gravesii* and *Q. hypoleucoides* groups, the least admixed individuals (Q values > 0.9 at $K = 4$) in each cluster according to our STRUCTURE analysis in dataset 2 were selected (see “Hybrid vs. Parental Comparisons” column in table A1). Welch’s t -test was used to determine whether H_O values were significantly higher in the *Q. tardifolia* group than in either parental group.

The R package triangularR (ver. 0.0.1; Wiens et al. 2025) was used to visualize the relationship between interclass heterozygosity (the proportion of loci heterozygous for ancestry between modeled parental species) and hybrid index (the proportion of alleles inherited from a single modeled parental species) strictly using ancestry informative markers (AIMs), alleles with large frequency differences between the putative parental taxa (here, *Q. gravesii* and *Q. hypoleucoides*). The same individuals included in the t -test were used for this analysis, and three allele frequency difference thresholds (0.75, 0.9, and 1.0) were explored.

Designation of hybrid classes. While the clustering approaches described above (STRUCTURE and PCA) are capable of identifying admixed individuals, they are ill suited for distinguishing hybrid classes based on admixture patterns, such as those shown by F_2 hybrids or more complex backcrosses (Anderson 2008). Therefore, to better characterize the hybrid nature of *Q. tardifolia* individuals and to complement our heterozygosity comparisons, we used the NewHybrids software (ver. 1.1 Beta 3; Anderson and Thompson 2002) to estimate posterior probabilities of individual assignment to different hybrid classes. We ran NewHybrids using the same set of individuals included in our heterozygosity comparisons (see table A1) and using the 211 loci identified by triangularR to be fixed for differences (allele frequency difference thresholds of 1.0) between the *Q. gravesii* and *Q. hypoleucoides* groups. We specified a 50,000 iteration burn-in and 500,000 sweeps and utilized the six default genotype frequency classes (corresponding to pure *Q. gravesii*, pure *Q. hypoleucoides*, F_1 hybrids, F_2 hybrids, backcrosses with *Q. gravesii*, and backcrosses with *Q. hypoleucoides*).

Polyploidy assessment. Because the initial results of our clustering and hybridization analyses indicated *Q. tardifolia* being composed exclusively of F_1 hybrids, we were motivated to investigate the ploidy levels in our dataset, to determine whether polyploidy in these hybrid individuals could be contributing to reproductive isolation. The ploidy status of different taxa was assessed with two approaches using different sample sets: flow cytometric analyses and bioinformatic assessments of SNP data (see below).

Polyploidy assessment—flow cytometric analyses. Most of the samples used in the flow cytometric analyses were distinct from those used for bioinformatic analyses; sample information can be found in table S4 [tables S1–S4 are available online]. At least three distinct plants were sampled for each species of interest (*Q. tardifolia*, $n = 5$; *Q. hypoleucoides*, $n = 3$; *Q. gravesii*, $n = 3$). Plant samples were prepared by co-chopping 2 cm² of *Quercus* fresh leaf or bud tissue with an equivalent amount of leaf tissue from the reference standard *Pisum sativum* “Ctirad” with

a double-edged razor blade in a petri dish containing 400 μL of extraction buffer (CyStain Precise P, Sysmex Corporation). Samples were then filtered through 50- μm nylon mesh into a cuvette, and 1600 μL of DAPI staining buffer (CyStain Precise P, Sysmex Corporation) was added. Samples were then processed for relative genome size estimates using a Quantum P Flow cytometer (Quantum Analysis, Münster, Germany). Each sample was repeated four times ($n = 4$). Relative genome size calculations were made for ploidy estimation by using DAPI and the reference standard *Pisum sativum* “Ctirad” (2C-value = 8.66 pg; Nix et al. 2024). Relative genome sizes were compared to measurements from *Q. rubra* because of the existing evidence of diploidy (estimated relative 2C-value = 1.95 pg; Kapoor et al. 2023) and relative genome size comparisons (Bai et al. 2012). Relative size estimates for unknowns observed between 1.80 and 2.00 pg were estimated to be diploids.

Polyploidy assessment—bioinformatic analyses. The software nQuire (Weiß et al. 2018) was used to estimate the likelihood of different ploidy models for each individual given their allele frequency distributions at variant sites with only two segregating bases. Compressed alignment files (in the .bam format) were converted into binary files using the nQuire create command and the default filtering parameters and minimum coverage threshold of 20 $\times$. Binary files were then denoised and passed along to the nQuire lrdmodel command to generate delta maximized log likelihood values for each ploidy scenario in each sample.

Results

Genotyping

After sequencing and demultiplexing, an average of 7.3 million (M) reads per individual (minimum, 46,516; maximum, 22.6 M) were obtained. After filtering and alignment to the *Quercus rubra* reference genome, an average of 3.0 M reads per individual (minimum, 13,813; maximum, 7.6 M) were retained for further analyses (table A1), with an average sequencing coverage across samples of $37.6 \times \pm 14.4 \times$ for the RADseq loci. Because of low numbers of reads (<1 M), four samples were excluded from most downstream analyses: two *Q. tardifolia* herbarium specimens (SYST-MOR-0006655 and SYST-MOR-0006657) and two *Q. miquihuaensis* samples (SYST-MOR-0007336 and SYST-MOR-0007341; see table A1). However, these were included in the initial STRUCTURE run of all samples and in the bioinformatic analyses of polyploidy (see “Clustering” and “Polyploidy Assessment”).

Clustering

STRUCTURE. Initial STRUCTURE runs including all samples (dataset 1; figs. S3, S4) indicate admixture in numerous individuals, including all individuals identified as *Q. tardifolia*, at the optimal K values (Evanno, $K = 4$; maximum likelihood, $K = 6$; see table S1). A sample set with more balanced representation of different taxa (dataset 2) highlights signals of hybridization in *Q. tardifolia* (figs. S5, S6), especially at the Evanno optimal K value of 4 (fig. 2), which we found to be the minimal K value that adequately represents divisions seen at higher K values. Other taxa (such as *Q. graciliformis*, *Q. robusta*, and *Q.*

canbyi) also show signs of admixture in this clustering analysis, although future analyses with more focused sampling of these groups would be ideal to verify these characterizations. Our sample set focused on *Q. tardifolia* and the indicated parental species (dataset 3) reinforces signals of hybridization seen in larger STRUCTURE runs, with *Q. tardifolia* individuals appearing as F_1 hybrids (approximately 50:50 ancestry proportions) between clusters composed of *Q. gravesii* and *Q. hypoleucoides* (fig. S7).

Principal component analysis and F_{ST} . Clustering patterns in our PCA of dataset 2 (figs. 3, S8 for version including all 84 samples) largely reinforce those shown in our STRUCTURE analyses. *Quercus tardifolia* clusters with known *Q. gravesii* $\times$ *hypoleucoides* hybrids, separated between *Q. gravesii* and *Q. hypoleucoides* clusters along PC axis 1, while clusters of *Q. hypoleucoides* and *Q. emoryi* are separated along PC axis 2. Additionally, *Q. graciliformis* individuals are distributed between clusters composed of *Q. gravesii* and *Q. emoryi* along PC axis 1, supporting the admixture patterns seen within this group in figure 2. Similarly, our PCA of dataset 3 shows *Q. tardifolia* positioned between clusters composed of *Q. gravesii* and *Q. hypoleucoides*, which are separated along PC axis 1 (fig. S9). These results, along with patterns of F_1 hybrid status for *Q. tardifolia* individuals in our STRUCTURE analyses, motivated our exploration of the hybrid nature of *Q. tardifolia* (see “Hybridization Analyses”). In correspondence with our clustering results, F_{ST} values indicate *Q. tardifolia* as being only weakly differentiated from clusters composed of *Q. gravesii* and *Q. hypoleucoides*, which are moderately differentiated from each other and from *Q. emoryi* (fig. S10).

Hybridization Analyses

Phylogenetic network. Reflecting results from our clustering analyses, the optimal phylogenetic network modeled by SNaQ (maximum one hybridization event allowed; negative log pseudolikelihood 2079.60) represents *Q. tardifolia* as a hybrid lineage (fig. 4). The network modeled using zero hybridization events had poorer fit (negative log pseudolikelihood 7849.03); when a maximum of two hybridization events was allowed, a network with an identical topology, hybrid edge position, and pseudolikelihood score was generated (data not shown). In the optimal network, the modeled hybrid edge (blue arrow) runs from an ancestral population of *Q. gravesii* to the *Q. tardifolia* group, which is sister to *Q. hypoleucoides*, with the proportion of *Q. gravesii* ancestry estimated to be 45.28%.

Heterozygosity comparisons. Welch’s t -tests confirm that mean heterozygosity is significantly higher in the group made up of *Q. tardifolia* (mean $H_O = 0.15$) than in the group composed of *Q. gravesii* (mean = 0.12) and the group composed of *Q. hypoleucoides* (mean = 0.08; see table S2). Triangle plots illustrate similar patterns of F_1 hybridization, with *Q. tardifolia* individuals showing higher interclass heterozygosity than parental groups regardless of the allele frequency threshold used (figs. 5, S11, S12).

Designation of hybrid classes. For all individuals, the maximal probability for hybrid class assignment is estimated by NewHybrids, and all *Q. tardifolia*/*Q. aff. tardifolia* individuals are estimated to be F_1 hybrids with the exception of SYST-MOR-0007272, which is estimated to be a backcross with *Q. hypoleucoides* (table S3). The assignment of this individual to

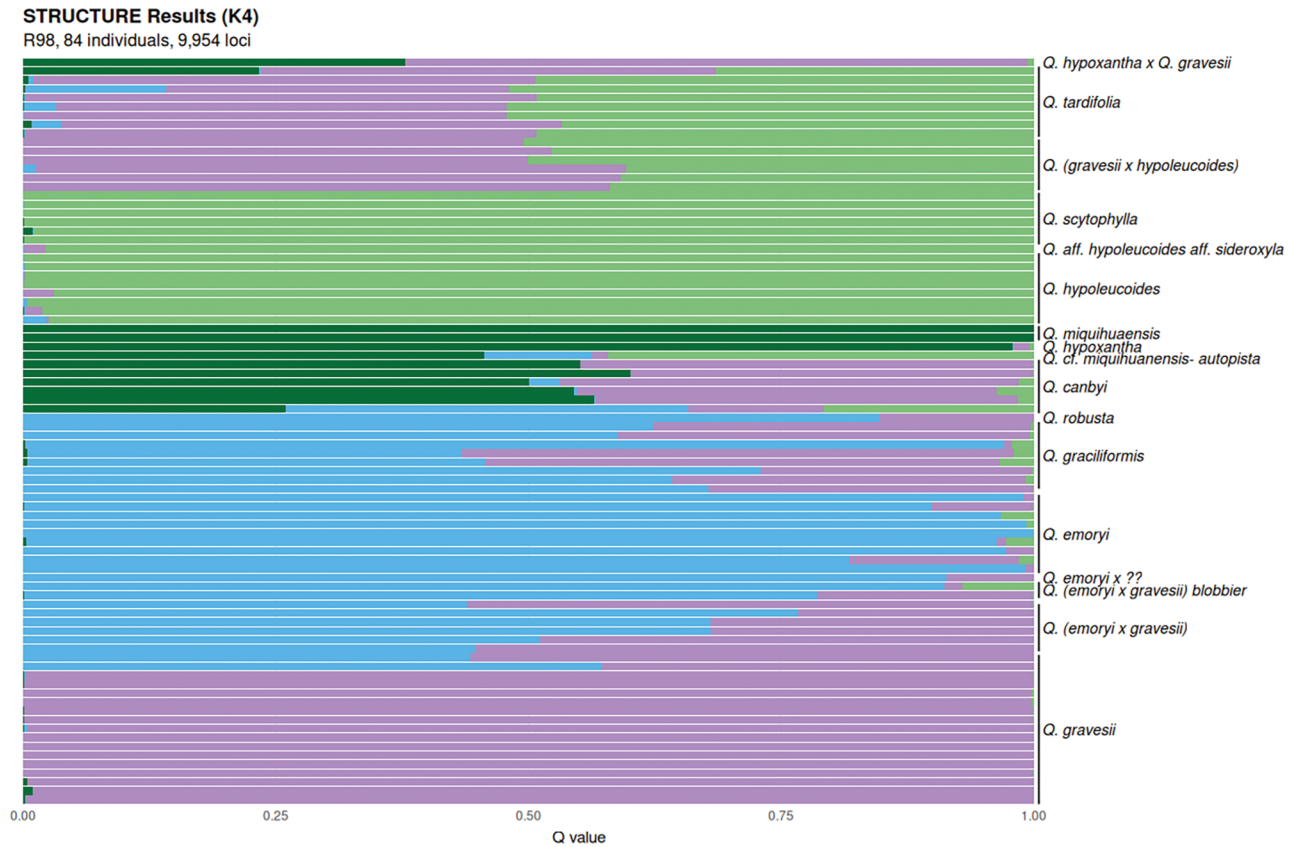


Fig. 2 STRUCTURE plot of 84 individuals assessed in STRUCTURE dataset 2. Horizontal bars represent individuals, and the different colors represent the genomic proportions from ancestral populations (determined by the K value). Classifications and abbreviated sample names are shown along the right-hand side. K values from 2 to 7 were explored; $K = 4$ is the optimal K value using the Evanno method (table S1), and adequately illustrates all the divisions seen between groups at higher values of K . STRUCTURE results for all K values are shown in figures S5 and S6.

the *Q. hypoleucoides* backcross hybrid class corresponds to results in our heterozygosity comparisons, which show SYSTMOR-0007272 having a larger hybrid index (closer to values corresponding to the *Q. hypoleucoides* group) and lower inter-class heterozygosity than the other *Q. tardifolia* individuals.

Polyploidy assessment. Mean 2C-values for *Q. tardifolia* range from 1.89 to 2.00 pg (table S4). *Quercus hypoleucoides* measurements range from 1.88 to 1.94 pg and *Q. gravesii* from 1.91 to 2.00 pg. All individuals included in these flow cytometric analyses were estimated as diploids with mean 2C relative genome sizes within the diploid range of 1.80 to 2.00 pg (minimum = 1.88 pg; maximum = 2.00 pg).

Using nQuire, 93.98% (156/166) of samples were modeled as diploid, including all seven of the wild-collected *Q. tardifolia* and *Q. aff. tardifolia* individuals, two herbarium specimens, and all seven known *Q. gravesii* × *hypoleucoides* hybrids (table A2). Of the 10 individuals not modeled as diploids, four have relatively low mean weighted coverage (<10×): two of the herbarium specimens (one *Q. tardifolia* and one *Q. hypoxantha* × *gravesii*) and two of the *Q. miquihuaensis* individuals. Of the six remaining nondiploid samples, three (one *Q. hypoxantha* and two *Q. emoryi*) were modeled as triploids, and three (one *Q. scytophylla* and two *Q. emoryi*) were modeled as tetraploids. These individ-

uals tended to have lower differences in the delta maximum log likelihood values between their most and least supported ploidy models, suggesting that our nQuire results are less reliable for these samples.

Discussion

The results of separate analyses—population genetic clustering approaches, phylogenetic network analysis, and comparisons of heterozygosity—all support the hypothesis that the specimens of the rare oak *Quercus tardifolia* sampled in this study are homoploid hybrids between *Q. gravesii* and *Q. hypoleucoides* or close relatives. A hybrid origin for *Q. tardifolia* has been hypothesized in the past (Jensen 1997), and Nixon and Muller (1993, p. 151) note that “specimens that are morphologically intermediate between *Q. coahuilensis* and *Q. gravesii* have been collected in the Chisos Mountains of Texas (*Q. tardifolia* C. H. Muller).” While *Q. coahuilensis* was unavailable for this study, it is morphologically similar to *Q. hypoxantha* and *Q. miquihuaensis* and is purported to be closely related to them (Nixon and Muller 1993). If *Q. tardifolia* were in fact a hybrid involving *Q. coahuilensis*, we would expect to detect *Q.*

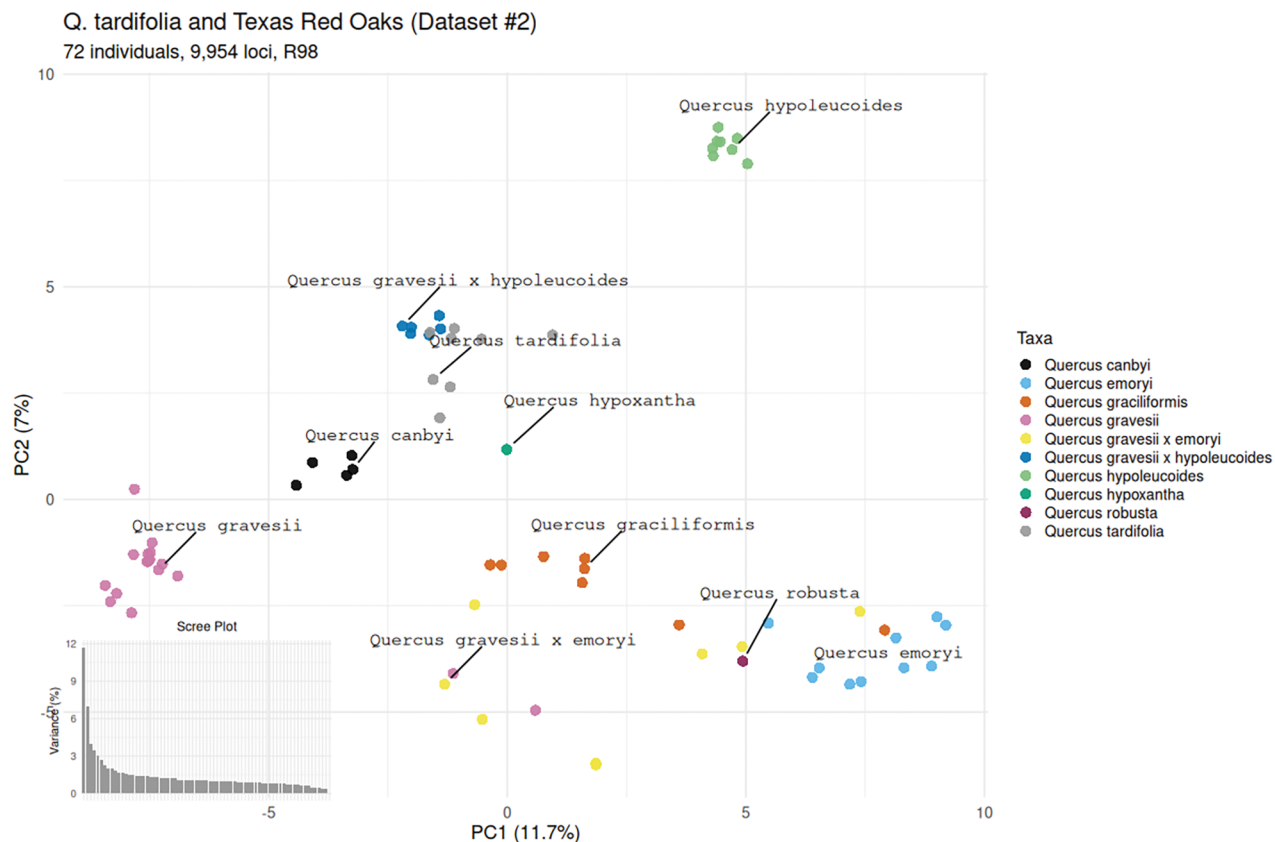


Fig. 3 Principal component analysis (PCA) of individuals included in dataset 2 of the clustering analyses. *Quercus scotyphylla*, *Q. miquihuaensis*, and several hybrid individuals were removed from this graphic to improve the clarity of the relationships between the most abundant taxa; see figure S8 for a PCA of all 84 individuals included in dataset 2. A scree plot depicting the proportion of variance explained by each principal component is shown in the bottom left. Each point represents an individual, and the different colors represent different taxa determined according to identifications (shown in the legend on the right); taxa group names are simplified for legibility (e.g., *Q. tardifolia* and *Q. aff. tardifolia* are combined into *tardifolia*). PC axes correspond to combinations of loci explaining the greatest variance within the data, such that points next to one another are more genetically similar.

tardifolia as being admixed with *Q. hypoxantha*, *Q. miquihuaensis*, or both. However, our results indicate that the parentage of *Q. tardifolia* does not include any of these species. Our findings provide the first genetic evidence for the origins of *Q. tardifolia*, and we consider the implications both for understanding the relationships between west Texas red oaks and for conserving the rare oak taxa in this part of the world.

Hybridization Patterns and Relationships between West Texas Red Oaks

Our study was motivated by the need to clarify the relations within this group of oaks, and for the most dominant taxa on the landscape, such as *Q. gravesii*, *Q. hypoleucoides*, and *Q. emoryi*, our results largely reflect the phylogenetic structure previously inferred for *Quercus* sect. *Lobatae* (Hipp et al. 2020), with *Q. gravesii* clustering further away from clusters composed of *Q. hypoleucoides* and *Q. emoryi*. *Quercus tardifolia* consistently emerges as a hybrid, and our polyploidy assessments illustrate that *Q. tardifolia* does not

constitute a polyploid species. While homoploid hybrid speciation is generally considered rare in plants, it is more common between closely related parents than distantly related ones (Soltis and Soltis 2009), and recent hybridization in *Quercus* is reported only between species in the same section (Stull et al. 2023). In spite of an exceptionally long history of hybridization in the genus (Zhou et al. 2022), there is only one documented homoploid hybrid species in *Quercus* (*Q. afares*; Mir et al. 2006), and the hybrid status of even that species has been questioned (Denk et al. 2023). Moreover, despite the fact that rare polyploids may be found in *Quercus* (Burda and Shchepotiev 1973; Butorina 1993), no polyploid species are known in the genus. *Quercus graciliformis*, *Q. robusta*, and *Q. canbyi* are also indicated as hybrids in our clustering analyses, although a more detailed assessment with a more thorough sampling effort is required to confirm this result; nevertheless, our findings may help explain the taxonomic confusions surrounding some of these names.

Our clustering analyses group *Q. tardifolia* individuals with the seven recognized *Q. gravesii* × *hypoleucoides* samples included in our dataset (syn. *Q. × inconstans*); a more

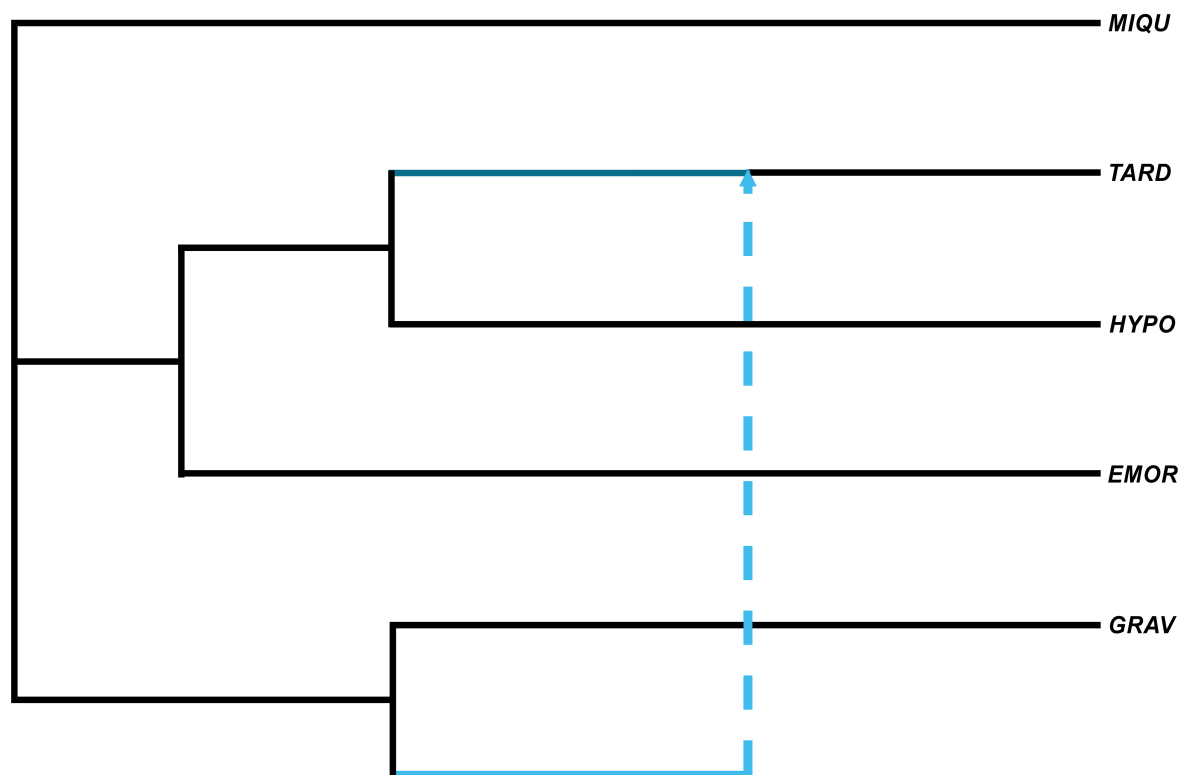


Fig. 4 Phylogenetic network constructed using the Species Networks applying Quartets (SNaQ) program, with a maximum of one hybridization event specified (negative log pseudolikelihood 2079.60). The dotted blue line represents the modeled hybrid edge, as estimated by SNaQ. Here, the placement of the hybrid edge indicates *Quercus tardifolia* as a hybrid sister to *Q. hypoleucoides* and with 45.28% of its genetic material from lineage ancestral of *Q. gravesii*. Refer to table A1 to see samples included in each group represented in this network.

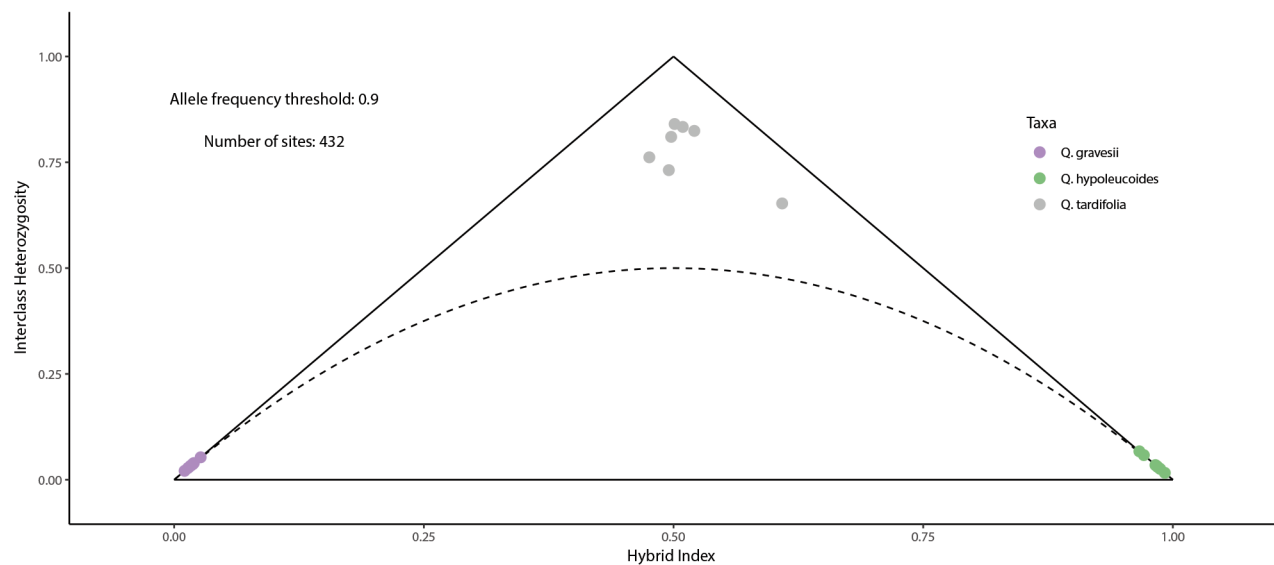


Fig. 5 Triangle plot generated by the triangulaR package indicating the interclass heterozygosity and hybrid index of *Quercus tardifolia*, *Q. gravesii*, and *Q. hypoleucoides* individuals (seven per group), using an allele frequency threshold of 0.9 (432 markers). The bottom right *Q. tardifolia* individual is SYST-MOR-0007272. Results using allele frequency thresholds of 0.75 and 1.0 are shown in figures S11 and S12, respectively.

detailed review of the morphological differences between these two entities is required to determine whether a continued taxonomic distinction between them is warranted (see fig. S2). Additionally, our clustering analyses consistently grouped the taxonomically distinct entities of *Q. hypoleucoides* and *Q. scytophylla* together. However, rather than indicating genuine shared identity, this result likely stems from the absence of informative loci needed to resolve differences between these two groups when analyzed within the wider context of other species (i.e., hierarchical structure, a common challenge in such clustering approaches). Additional STRUCTURE analyses that included only *Q. hypoleucoides* and *Q. scytophylla* individuals distinguished the two (data not shown).

Our hybridization findings raise additional questions regarding the origin of *Q. tardifolia* and its relation to other taxa on the landscape. It is noteworthy that the majority of *Q. tardifolia* samples, regardless of their location, show similar patterns of F_1 hybridization, with approximately 50% ancestry coming from each parental species (figs. 2, 4), because while *Q. hypoleucoides* is known to occur in the Davis Mountains, it is not known to occur in the Chisos (fig. 1). The closest-known *Q. hypoleucoides* occurrences to the Chisos would be in the Sierra del Carmen in Coahuila, Mexico, approximately 40 miles away (Wood et al. 1999). However, one individual identified as *Q. aff. scytophylla* aff. *hypoleucoides* (SYSTMOR-0006692) was located in the Chisos Mountains; whether this explains the genetic signatures of *Q. hypoleucoides* in *Q. tardifolia* individuals sampled from the Chisos requires a more focused analysis. The consistent ancestry proportions across sampling localities are especially striking given our identification of a single backcrossed individual, SYSTMOR-0007272, which is located in the Davis Mountains and is proximate to other *Q. hypoleucoides* and *Q. tardifolia* individuals. The presence of this backcrossed individual implies that *Q. tardifolia* hybrids are fertile and raises the question of why there is a relative absence of patterns of more advanced introgression among the individuals observed in the field and in our dataset. There are several possible explanations for this trend. First, our sampling could have been too limited to include more backcrossed individuals; however, we sampled extensively and endeavored to collect a wide range of morphological oddities in the field, which should have increased our chances of finding introgressed individuals. Second, hybrids may have phenological or ecological differences that make backcrossing unlikely. This is possible; we were unable to assess phenology in the field, though ecologically the intermediates were often codistributed with *Q. gravesii*. Finally, hybrids could be less fit than both parents, leading to maintained zones of recurrently formed F_1 hybrids (Abbott 2017), although we have no data to demonstrate this. Establishing which of these scenarios is the case, or whether any individuals are backcrossed to *Q. gravesii*, requires a clearer understanding of the extent of reproductive isolation between *Q. tardifolia* and either parental species and perhaps more expansive sampling.

Ancestry proportions in hybrid individuals are typically expected to be impacted by the relative abundance and geographic proximity of parental species (Lepais et al. 2009; Zeng et al. 2011; Klein et al. 2017), at least historically (Maze 1968; McVay et al. 2017a, 2017b), and the results of our clustering analyses correspond to this trend, with admixed individuals being represented as intermediates between clusters composed

of the most abundant species on the landscape. For instance, figure 2 illustrates *Q. graciliformis* individuals as being admixed between clusters composed of *Q. gravesii* and *Q. emoryi*, a trend also reflected in our PCA. More refined analyses focused on *Q. graciliformis* and potentially interbreeding species, including the type specimen (which was not included in our sampling), would help clarify the admixture patterns observed here.

Caveats

Despite being widely used for delimiting populations and species, it is possible for STRUCTURE and similar clustering approaches to represent similarities due to ancient shared ancestry between individuals as more recent genetic admixture, leading to misinterpretation of resulting bar plots (Lawson et al. 2018). To account for this in our study, we validated our STRUCTURE results by assessing results across K values (Gilbert et al. 2012), controlling for uneven sampling (Puechmaile 2016), and comparing our results with those from other analyses (PCA, SNaQ, heterozygosity comparisons, and hybrid class predictions from NewHybrids). Our recognition of the *Q. tardifolia* individuals in this study as homoploid hybrids is made more robust by their consistent characterization across multiple approaches.

While our analyses consistently reflect patterns of hybridization in *Q. tardifolia*, aspects of our methodology limit our ability to generalize this result across the taxon as it is recognized more broadly. Our sampling of *Q. tardifolia* consists of only seven individuals from the Davis and Chisos Mountains, which at the time of sampling were all known individuals. More recent field expeditions have identified *Q. tardifolia* individuals in more areas than previously recognized, meaning that the results presented here are based on a limited number of the locations in which *Q. tardifolia* is found. Additionally, we were unable to assess the genetic makeup (and hybridization status) of the type and other historical specimens of *Q. tardifolia* because of poor sequencing results of herbarium samples (which ranged between 28 and 93 yr old), and across all taxa, our conclusions are limited to wild-collected specimens from 2011 to 2022. Approaches for assessing historic DNA could be valuable for determining whether the clustering patterns shown here are supported or refuted by herbarium specimens.

Conservation Implications and Next Steps

Quercus tardifolia has previously been assessed as a species of concern (Beckman et al. 2019; “Critically Imperiled,” NatureServe, https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.129171/Quercus_tardifolia) but our work raises the question as to whether it should be recognized as a species at all. Some important conservation implications of our results, however, matter whatever the resolution of the taxonomic question. While hybrid individuals and hybrid-derived populations have often been disregarded as unimportant for the preservation of a species, the advent of increasingly accessible sequencing technologies has led to a heightened appreciation of both the frequency of hybridization in nature and its potential to contribute to local adaptation in wild populations (Allendorf et al. 2001). In the United States, the federal protection of hybrids is handled on a case-by-case basis under the Endangered Species

Act, with little guidance regarding the circumstances under which hybrids should be protected (although see Wayne and Shaffer 2016). Our results indicating hybrid fertility (as evidenced by the presence of a backcrossed individual), as well as our relatively limited sampling, suggest that more information is needed in order to establish the urgency and value of any conservation action. Conservation status assessments under both the IUCN Red List and NatureServe criteria are often not conducted for hybrids (IUCN, <https://www.iucnredlist.org/about/faqs>; NatureServe, https://help.natureserve.org/biotics/content/record_management/Element_Files/Element_Tracking/ETRACK_Definitions_of_Heritage_Conservation_Status_Ranks.htm?) but can be applied. In this and similar scenarios, the key question is whether the entity in question is worthy of conservation attention or whether it is a natural spontaneous hybrid. While more information is needed to answer this question, the results of this study have provided an initial understanding of the origins of *Q. tardifolia*, demonstrating the value that genetic data can add to conservation assessments.

Regardless of its conservation status, the hybrid status of *Q. tardifolia* may be contributing to the persistence of parental species on the landscape. Hybridization can serve as either a threat (via processes such as genetic swamping) or an aid to the conservation of parental species (Quilodr  n et al. 2020). In the Brazilian Cerrado forest, for example, the legally protected tree *Dimorphandra wilsonii* was found to be a hybrid between two sympatric species and may be acting as a source of genetic diversity for one of the parentals (Muniz et al. 2020). There are many documented examples of hybrid plant species serving valuable roles within the ecosystems in which they arise. In a system composed of three *Populus* species, hybrids have been shown to affect the geographic distributions of gall-forming arthropods (Floate et al. 2016), supporting the hypothesis that the presence of hybrid genotypes may reduce selective pressures associated with host shifting (Floate and Whitham 1993). Changes in the leaf chemistry of hybrid individuals can have ecosystem-wide consequences due to impacts in decomposition rates (Stange et al. 2021), and *Helianthus* hybrids can withstand extreme environments uninhabitable by parental species (Rieseberg et al. 2003), allowing for adaption to new ecological conditions. Adaptive introgression appears to have played a role in the expansion of at least some oaks into novel environments (Leroy et al. 2020; O'Donnell et al. 2021) and may play an increasingly

important role in stressful environments as climates change. The environmental niches of *Q. hypoleucoides* and *Q. gravesii* largely overlap, but *Q. hypoleucoides* has a broader geographic range extending to the Sierra Madre Occidental, while *Q. gravesii* is more restricted, with a range extending south into the Sierra Madre Oriental (fig. 1). Clarifying how *Q. tardifolia* fits into the larger ecological context of the Trans-Pecos region, and whether it serves any unique roles within that context, can help to determine what conservation interventions for this rare oak, if any, would be warranted.

Beyond more extensive spatial sampling and comparison with historic specimens, the hybrid nature of *Q. tardifolia* could be assessed by conducting controlled crosses between *Q. gravesii* and *Q. hypoleucoides* and testing the viability of resulting acorns. Additionally, phenological observations, like flowering time for each species, could be valuable for understanding how these hybrid individuals form in the wild, especially given the absence of *Q. hypoleucoides* in the Chisos Mountains. Finally, observations of *Q. tardifolia* and its biotic interactions within the Trans-Pecos may help to elucidate its ecological role, if any. While our analyses have helped to clarify the origin of *Q. tardifolia* and its relationships with sympatric species, further genetic analyses are needed to more fully understand the nature of the other rare taxa in this region of west Texas and to ensure their protection and survival in a changing environment.

Acknowledgments

This project work was done in collaboration with the US Botanic Garden. We thank the US National Park Service, namely, Carolyn Whiting, Phillip Varela, Herbert Young, Deb Manley, and Stephen Lantz, for their assistance with research permits and field support in Big Bend National Park. From the Nature Conservancy we thank Charlotte Reemts and Kaylee French for their assistance with permits and access to the Davis Mountains Preserve. From Sul Ross State University A. Michael Powell Herbarium (SRSC), we thank A. Michael Powell and Shirley Powell for their generous assistance with specimen access and their decades of local knowledge that helped guide our efforts in many ways. We finally extend our gratitude to the late Patricia Manning for invaluable guidance on field sites, stemming from her years of botanical explorations in west Texas.

Literature Cited

- Abbott RJ 2017 Plant speciation across environmental gradients and the occurrence and nature of hybrid zones: hybrid zones and plant speciation. *J Syst Evol* 55:238–258. <https://doi.org/10.1111/jse.12267>.
- Alexander DH, J Novembre, K Lange 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Alix K, PR G  rard, T Schwarzacher, JSP Heslop-Harrison 2017 Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann Bot* 120:183–194. <https://doi.org/10.1093/aob/mcx079>.
- Allendorf FW, RF Leary, P Spruell, JK Wenburg 2001 The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol* 16:613–622. [https://doi.org/10.1016/s0169-5347\(01\)02290-x](https://doi.org/10.1016/s0169-5347(01)02290-x).
- Anderson EC 2008 Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philos Trans R Soc B* 363:2841–2850. <https://doi.org/10.1098/rstb.2008.0043>.
- Anderson EC, EA Thompson 2002 A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217–1229. <https://doi.org/10.1093/genetics/160.3.1217>.
- Andrews KR, JM Good, MR Miller, G Luikart, PA Hohenlohe 2016 Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17:81–92. <https://doi.org/10.1038/nrg.2015.28>.
- Andrews S 2010 FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

- Bai C, WS Alverson, A Follansbee, DM Waller 2012 New reports of nuclear DNA content for 407 vascular plant taxa from the United States. *Ann Bot* 110:1623–1629. <https://doi.org/10.1093/aob/mcs222>.
- Baird NA, PD Etter, TS Atwood, MC Currey, AL Shiver, ZA Lewis, EU Selker, WA Cresko, EA Johnson 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376. <https://doi.org/10.1371/journal.pone.0003376>.
- Barton N, G Hewitt 1985 Analysis of hybrid zones. *Annu Rev Ecol Evol Syst* 16:113–148. <https://doi.org/10.1146/ANNUREV.ES.16.110185.000553>.
- Beckman E, A Meyer, A Denvir, D Gill, G Man, D Pivorunas, K Shaw, M Westwood 2019 Conservation gap analysis of native U.S. oaks.” Morton Arboretum. https://mortonarb.org/app/uploads/2021/05/conservation-gap-analysis-of-native-US-oaks_sm.pdf.
- Burda RI, FL Shchepotiev 1973 Spontaneous polyploidy in seedlings of multi-seeded acorns of *Quercus robur* L. *Cytol Genet* 7:140–143.
- Burger WC 1975 The species concept in *Quercus*. *Taxon* 24:45–50. <https://doi.org/10.2307/1218998>.
- Butorina AK 1993 Cytogenetic study of diploid and spontaneous triploid oaks, *Quercus robur* L. *Ann Sci For* 50(suppl):144s–150s. <https://doi.org/10.1051/forest:19930714>.
- Carrero C, D Jerome, E Beckman, A Byrne, AJ Coombes, M Deng, A González Rodríguez, et al 2020 The Red List of Oaks 2020. Morton Arboretum. <https://careyrussell.com/s/RedListOaks2020.pdf>.
- Carstens BC, TA Pelletier, NM Reid, JD Satler 2013 How to fail at species delimitation. *Mol Ecol* 22:4369–4383. <https://doi.org/10.1111/mec.12413>.
- Cavender-Bares J 2019 Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytol* 221:669–692. <https://doi.org/10.1111/nph.15450>.
- Cavender-Bares JM, E Nelson, JE Meireles, JR Lasky, DA Miteva, DJ Nowak, WD Pearse, et al 2022 The hidden value of trees: quantifying the ecosystem services of tree lineages and their major threats across the contiguous US. *PLoS Sustain Transform* 1: e0000010. <https://doi.org/10.1371/journal.pstr.0000010>.
- Cowan R, M Baranski 1976 An analysis of variation within white oak (*Quercus alba* L.). *Taxon* 25:347. <https://doi.org/10.2307/1219471>.
- Curtu AL, O Gailing, R Finkeldey 2007 Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evol Biol* 7:218. <https://doi.org/10.1186/1471-2148-7-218>.
- Darwin C 1872 On the origin of species by means of natural selection. J Murray, London.
- Degnan JH 2018 Modeling hybridization under the network multispecies coalescent. *Syst Biol* 67:786–799. <https://doi.org/10.1093/sysbio/syy040>.
- Denk T, GW Grimm, AL Hipp, JM Bouchal, E-D Schulze, MC Simeone 2023 Niche evolution in a northern temperate tree lineage: biogeographical legacies in cork oaks (*Quercus* section *Cerris*). *Ann Bot* 131:769–787. <https://doi.org/10.1093/aob/mcad032>.
- De Queiroz K 2007 Species concepts and species delimitation. *Syst Biol* 56:879–886. <https://doi.org/10.1080/10635150701701083>.
- DeSalle R, G Amato 2004 The expansion of conservation genetics. *Nat Rev Genet* 5:702–712. <https://doi.org/10.1038/nrg1425>.
- Desmond SC, M Garner, S Flannery, AT Whittemore, AL Hipp 2021 Leaf shape and size variation in bur oaks: an empirical study and simulation of sampling strategies. *Am J Bot* 108:1540–1554. <https://doi.org/10.1002/ajb2.1705>.
- Dzialuk A, I Chybicki, M Welc, E Sliwinska, J Burczyk 2007 Presence of triploids among oak species. *Ann Bot* 99:959–964. <https://doi.org/10.1093/aob/mcm043>.
- Evanno G, S Regnaut, J Goudet 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Falush D, M Stephens, JK Pritchard 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587. <https://www.ncbi.nlm.nih.gov/pubmed/12930761>.
- Fitzpatrick BM 2012 Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evol Biol* 12:131. <https://doi.org/10.1186/1471-2148-12-131>.
- Floate KD, J Godbout, MK Lau, N Isabel, TG Whitham 2016 Plant-herbivore interactions in a trispecific hybrid swarm of *Populus*: assessing support for hypotheses of hybrid bridges, evolutionary novelty and genetic similarity. *New Phytol* 209:832–844. <https://doi.org/10.1111/nph.13622>.
- Floate KD, TG Whitham 1993 The “hybrid bridge” hypothesis: host shifting via plant hybrid swarms. *Am Nat* 141:651–662. <https://doi.org/10.1086/285497>.
- Frankel SJ, J Juzwik, DM Rizzo 2022 Forests (oaks) in North America. <https://research.fs.usda.gov/treesearch/67252>.
- Gilbert KJ, RL Andrew, DG Bock, MT Franklin, NC Kane, J-S Moore, BT Moyers, et al 2012 Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Mol Ecol* 21:4925–4930. <https://doi.org/10.1111/j.1365-294X.2012.05754.x>.
- Gruber B, PJ Unmack, OF Berry, A Georges 2018 DART: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* 18:691–699. <https://doi.org/10.1111/1755-0998.12745>.
- Harrison RG 1986 Pattern and process in a narrow hybrid zone. *Heredity* 56:337–349. <https://doi.org/10.1038/hdy.1986.55>.
- Hausdorf B 2025 Species delimitation using genomic data: options and limitations. *Mol Ecol* 34:e17717. <https://doi.org/10.1111/mec.17717>.
- Hipp AL, PS Manos, M Hahn, M Avishai, C Bodénès, J Cavender-Bares, AA Crowl, et al 2020 Genomic landscape of the global oak phylogeny. *New Phytol* 226:1198–1212. <https://doi.org/10.1111/nph.16162>.
- Hipp AL, JA Weber 2008 Taxonomy of Hill’s oak (*Quercus ellipsoidalis*: Fagaceae): evidence from AFLP data. *Syst Bot* 33:148–158. <https://doi.org/10.1600/036364408783887320>.
- Hodel RGJ, S Chen, AC Payton, SF McDaniel, P Soltis, DE Soltis 2017 Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing micro-satellites and RAD-Seq and investigating loci filtering. *Sci Rep* 7:17598. <https://doi.org/10.1038/s41598-017-16810-7>.
- Huang J-P 2020 Is population subdivision different from speciation? from phylogeography to species delimitation. *Ecol Evol* 36:303. <https://doi.org/10.1002/ece3.6524>.
- Jakobsson M, NA Rosenberg 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>.
- Jensen RJ 1997 *Quercus* Linnaeus sect. *Lobatae* Loudon. Pp. 447–468 in *Flora of North America* Editorial Committee, eds. *Flora of North America north of Mexico*. Oxford University Press, New York.
- Jombart T 2008 adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>.
- Kapoor B, J Jenkins, J Schmutz, T Zhebentyayeva, C Kuelheim, M Coggeshall, C Heim, et al 2023 A haplotype-resolved chromosome-scale genome for *Quercus rubra* L. provides insights into the genetics of adaptive traits for red oak species. *G3* 13:jkad209. <https://doi.org/10.1093/g3journal/jkad209>.
- Klein EK, L Lagache-Navarro, RJ Petit 2017 Demographic and spatial determinants of hybridization rate. *J Ecol* 105:29–38. <https://doi.org/10.1111/1365-2745.12674>.

- Knapp WM, A Frances, R Noss, RFC Naczi, A Weakley, GD Gann, BG Baldwin, et al 2021 Vascular plant extinction in the continental United States and Canada. *Conserv Biol* 35:360–368. <https://doi.org/10.1111/cobi.13621>.
- Kopelman NM, J Mayzel, M Jakobsson, NA Rosenberg, I Mayrose 2015 CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>.
- Kremer A, AL Hipp 2020 Oaks: an evolutionary success story. *New Phytol* 226:987–1011. <https://doi.org/10.1111/nph.16274>.
- Lawson DJ, L van Dorp, D Falush 2018 A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun* 9:3258. <https://doi.org/10.1038/s41467-018-05257-7>.
- Lepais O, RJ Petit, E Guichoux, JE Lavabre, F Alberto, A Kremer, S Gerber 2009 Species relative abundance and direction of introgression in oaks. *Mol Ecol* 18:2228–2242. <https://doi.org/10.1111/j.1365-294X.2009.04137.x>.
- Leroy T, C Plomion, A Kremer 2020 Oak symbolism in the light of genomics. *New Phytol* 226:1012–1017. <https://doi.org/10.1111/nph.15987>.
- Li H, R Durbin 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Little EL 1971 Atlas of United States trees. US. Department of Agriculture, Forest Service, Washington, DC.
- Ma J, CI Amos 2012 Principal components analysis of population admixture. *PLoS ONE* 7:e40115. <https://doi.org/10.1371/journal.pone.0040115>.
- MacDougal DT 1907 Hybridization of wild plants. *Bot Gaz* 43:45–58.
- Mace GM 2004 The role of taxonomy in species conservation. *Philos Trans R Soc B* 359:711–719. <https://doi.org/10.1098/rstb.2003.1454>.
- Mallet J 1995 A species definition for the modern synthesis. *Trends Ecol Evol* 10:294–299. [https://doi.org/10.1016/0169-5347\(95\)90031-4](https://doi.org/10.1016/0169-5347(95)90031-4).
- 2005 Hybridization as an invasion of the genome. *Trends Ecol Evol* 20:229–237. <https://doi.org/10.1016/j.tree.2005.02.010>.
- 2020 Alternative views of biological species: reproductively isolated units or genotypic clusters? *Natl Sci Rev* 7:1401–1407. <https://doi.org/10.1093/nsr/nwaa116>.
- Mastretta-Yanes A, N Arrigo, N Alvarez, TH Jorgensen, D Piñero, BC Emerson 2015 Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol Ecol Resour* 15:28–41. <https://doi.org/10.1111/1755-0998.12291>.
- Mayden RL 1997 A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381–423 in MF Claridge, HA Dawah, MR Wilson, eds. *Species: the units of diversity*. Chapman & Hall, London.
- Maze J 1968 Past hybridization between *Quercus macrocarpa* and *Quercus gambelii*. *Brittonia* 20:321. <https://doi.org/10.2307/2805689>.
- McVay JD, D Hauser, AL Hipp, PS Manos 2017a Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* 60:733–742. <https://doi.org/10.1139/gen-2016-0206>.
- McVay JD, AL Hipp, PS Manos 2017b A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc R Soc B* 284:20170300. <https://doi.org/10.1098/rspb.2017.0300>.
- Meirmans PG 2019 Subsampling reveals that unbalanced sampling affects STRUCTURE results in a multi-species dataset. *Heredity* 122:276–287. <https://doi.org/10.1038/s41437-018-0124-8>.
- Mir C, L Toumi, P Jarne, V Sarda, F Di Giusto, R Lumaret 2006 Endemic North African *Quercus afares* Pomel originates from hybridisation between two genetically very distant oak species (*Q. suber* L. and *Q. canariensis* Willd.): evidence from nuclear and cytoplasmic markers. *Heredity* 96:175–184. <https://doi.org/10.1038/sj.hdy.6800782>.
- Mohn RA, M Garner, PS Manos, AL Hipp 2026 Variant calling in the Goldilocks zone: how reference genome choice and read mapping stringency impact heterozygosity estimates and phylogenetic analyses. *Mol Ecol Resour* 26:e70079. <https://doi.org/10.1111/1755-0998.70079>.
- 1940 Oaks of Trans-Pecos Texas. *Am Midl Nat* 24:703. <https://doi.org/10.2307/2420868>.
- Muller CH 1951 The oaks of Texas. Vol 1. Texas Research Foundation, Renner.
- Muniz AC, JP Lemos-Filho, HA Souza, RC Marinho, RS Buzatti, M Heuertz, MB Lovato 2020 The protected tree *Dimorphandra wilsonii* (Fabaceae) is a population of inter-specific hybrids: recommendations for conservation in the Brazilian Cerrado/Atlantic Forest ecotone. *Ann Bot* 126:191–203. <https://doi.org/10.1093/aob/mcaa066>.
- Musich R, L Cadle-Davidson, MV Osier 2021 Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. *Front Plant Sci* 12:657240. <https://doi.org/10.3389/fpls.2021.657240>.
- Nix J, TG Ranney, NP Lynch, H Chen 2024 Flow cytometry for estimating plant genome size: revisiting assumptions, sources of variation, reference standards, and best practices. *J Am Soc Hortic Sci* 149:131–141.
- Nixon KC, CH Muller 1993 The *Quercus hypoxantha* complex (Fagaceae) in northeastern Mexico. *Brittonia* 45:146. <https://doi.org/10.2307/2807497>.
- Noss RF, JM Cartwright, D Estes, T Witsell, G Elliott, D Adams, M Albrecht, et al 2021 Improving species status assessments under the U.S. Endangered Species Act and implications for multispecies conservation challenges worldwide. *Conserv Biol* 35:1715–1724. <https://doi.org/10.1111/cobi.13777>.
- Novembre J, M Stephens 2008 Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40:646–649. <https://doi.org/10.1038/ng.139>.
- O'Donnell ST, ST Fitz-Gibbon, VL Sork 2021 Ancient introgression between distantly related white oaks (*Quercus* sect. *Quercus*) shows evidence of climate-associated asymmetric gene exchange. *J Hered* 112:663–670. <https://doi.org/10.1093/jhered/esab053>.
- Olave M, A Meyer 2020 Implementing large genomic single nucleotide polymorphism data sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of cichlid fish. *Syst Biol* 69:848–862. <https://doi.org/10.1093/sysbio/syaa005>.
- Palmer EJ 1948 Hybrid oaks of North America. *J Arnold Arbor Harv Univ* 29:1–48. <https://doi.org/10.5962/p.185588>.
- Parmesan C, MC Singer, B Wee, S Mikheyev 2023 The case for prioritizing ecology/behavior and hybridization over genomics/taxonomy and species' integrity in conservation under climate change. *Biol Conserv* 281:109967. <https://doi.org/10.1016/j.biocon.2023.109967>.
- Patterson N, AL Price, D Reich 2006 Population structure and eigenanalysis. *PLoS Genet* 2:e190. <https://doi.org/10.1371/journal.pgen.0020190>.
- Payseur BA, LH Rieseberg 2016 A genomic perspective on hybridization and speciation. *Mol Ecol* 25:2337–2360. <https://doi.org/10.1111/mec.13557>.
- Petit RJ, C Bodénès, A Ducouso, G Roussel, A Kremer 2004 Hybridization as a mechanism of invasion in oaks: research review. *New Phytol* 161:151–164. <https://doi.org/10.1046/j.1469-8137.2003.00944.x>.
- Petry W, S Taylor 2025 Wpetry/USTreeAtlas: V.1.1-Patch. Zenodo. <https://doi.org/10.5281/ZENODO.15468153>.
- Piett S, HA Hager, C Gerrard 2015 Characteristics for evaluating the conservation value of species hybrids. *Biodivers Conserv* 24:1931–1955. <https://doi.org/10.1007/s10531-015-0919-3>.

- Pritchard JK, M Stephens, P Donnelly 2000 Inference of population structure using multilocus genotype data. *Genetics* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>.
- Pritchard JK, X Wen, D Falush 2010 Documentation for STRUCTURE software. https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf.
- Puechmaille SJ 2016 The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol Ecol Resour* 16:608–627. <https://doi.org/10.1111/1755-0998.12512>.
- Quilodr  n CS, JI Montoya-Burgos, M Currat 2020 Harmonizing hybridization dissonance in conservation. *Comm Biol* 3:391. <https://doi.org/10.1038/s42003-020-1116-9>.
- Rannala B 2015 The art and science of species delimitation. *Curr Zool* 61:846–853. <https://doi.org/10.1093/czoolo/61.5.846>.
- R Core Team 2022 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Renault M 2021 The strange quest to save North America’s most elusive oak tree. *New Republic*. <https://newrepublic.com/article/164429/quest-elusive-oak-tree-biodiversity>.
- 2022 Found: one oak tree, famously missing. *New Republic*. <https://newrepublic.com/article/167008/found-one-oak-tree-famously-missing>.
- Rieseberg LH, O Raymond, DM Rosenthal, Z Lai, K Livingstone, T Nakazato, JL Durphy, AE Schwarzbach, LA Donovan, C Lexer 2003 Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216. <https://doi.org/10.1126/science.1086949>.
- Rivera-Col  n AG, J Catchen 2022 Population genomics analysis with RAD, reprised: Stacks 2. *Methods Mol Biol* 2498:99–149. https://doi.org/10.1007/978-1-0716-2313-8_7.
- Rosenberg NA, LM Li, R Ward, JK Pritchard 2003 Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73:1402–1422. <https://doi.org/10.1086/380416>.
- Rubio De Casas R, P Vargas, E P  rez-Corona, E Manrique, JR Quintana, C Garc  a-Verdugo, L Balaguer 2007 field patterns of leaf plasticity in adults of the long-lived evergreen *Quercus coccifera*. *Ann Bot* 100:325–334. <https://doi.org/10.1093/aob/mcm112>.
- Sax HJ 1930 Chromosome numbers in *Quercus*. *J Arnold Arbor Harv Univ* 11:220–223. <https://doi.org/10.5962/p.185215>.
- Sol  s-Lemus C, C An   2016 Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet* 12:e1005896. <https://doi.org/10.1371/journal.pgen.1005896>.
- Soltis PS, DE Soltis 2009 The role of hybridization in plant speciation. *Annu Rev Plant Biol* 60:561–588. <https://doi.org/10.1146/annurev.arplant.043008.092039>.
- Stange M, RDH Barrett, AP Hendry 2021 The importance of genomic variation for biodiversity, ecosystems and people. *Nat Rev Genet* 22:89–105. <https://doi.org/10.1038/s41576-020-00288-7>.
- Stanton DWG, P Frandsen, RK Waples, R Heller, I-RM Russo, PA Orozco-terWengel, C-E Tingskov Pedersen, HR Siegismund, MW Bruford 2019 More grist for the mill? species delimitation in the genomic era and its implications for conservation. *Conserv Genet* 20:101–113. <https://doi.org/10.1007/s10592-019-01149-5>.
- Stull GW, KK Pham, PS Soltis, DE Soltis 2023 Deep reticulation: the long legacy of hybridization in vascular plant evolution. *Plant J* 114:743–766. <https://doi.org/10.1111/tpj.16142>.
- Trelease W 1917 Naming American hybrid oaks. *Proc American Philos Soc* 56:44–52. <https://www.jstor.org/stable/983961?seq=1>.
- Tucker JM 1961 Studies in the *Quercus undulata* complex: a preliminary statement. *Am J Bot* 48:202–208. <https://doi.org/10.1002/j.1537-2197.1961.tb11626.x>.
- Wayne RK, HB Shaffer 2016 Hybridization and endangered species protection in the molecular era. *Mol Ecol* 25:2680–2689. <https://doi.org/10.1111/mec.13642>.
- Weir BS, CC Cockerham 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358–1370. <https://doi.org/10.2307/2408641>.
- Wei   CL, M Pais, LM Cano, S Kamoun, HA Burbano 2018 nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* 19:122. <https://doi.org/10.1186/s12859-018-2128-z>.
- Wen D, Y Yu, J Zhu, L Nakhleh 2018 Inferring phylogenetic networks using PhyloNet. *Syst Biol* 67:735–740. <https://doi.org/10.1093/sysbio/syy015>.
- Wiens BJ, LH DeCicco, JP Colella 2025 triangulaR: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones. *Heredity* 134:251–262. <https://doi.org/10.1038/s41437-025-00760-2>.
- Wood S, G Harper, E Muldavin, P Neville 1999 Vegetation map of the Sierra Del Carmen, U.S.A. and Mexico. <https://nhnm.unm.edu/system/files/2024-04/U99WOO02NMUS.pdf>.
- Zeng Y-F, W-J Liao, RJ Petit, D-Y Zhang 2011 Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Mol Ecol* 20:4995–5011. <https://doi.org/10.1111/j.1365-294X.2011.05354.x>.
- Zhou B-F, S Yuan, AA Crowl, Y-Y Liang, Y Shi, X-Y Chen, Q-Q An, M Kang, PS Manos, B Wang 2022 Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nat Comm* 13:1320. <https://doi.org/10.1038/s41467-022-28917-1>.